ACToR – Aggregated Computational Toxicology Resource

Release v2009Q3 (September 2009) Produced by the U.S. Environmental Protection Agency, National Center for Computational Toxicology

Table of Contents

Introduction

Overview: how to use ACToR

Data Collection

Data Collection View

Generic Chemical View

Search by Name

Search by CAS Number

Search by Structure

Browsing Assay

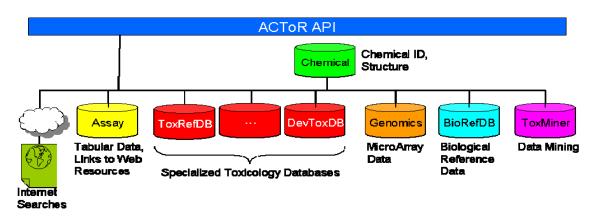
Assay View

Assay Data Page

Glossary

Introduction

ACTOR (Aggregated Computational Toxicology Resource) is a collection of databases collated or developed by the US EPA National Center for Computational Toxicology (NCCT). Over 400 sources of publicly available data on environmental chemicals have been brought together and made searchable by chemical names, <u>CAS</u> number, structure drawings, <u>InCHI</u> strings and <u>SMILES</u>. This database contains data such as <u>hazardous</u> testing results, chemical structure, physico-chemical values, <u>in vitro</u> assay data, <u>exposure data</u>, and <u>in vivo</u> toxicology data on hundreds of thousands of generic chemicals. These chemicals include, but are not limited to, <u>high and medium production volume industrial chemicals</u>, pesticides (<u>active</u> and <u>inert</u> ingredients), food additives and potential ground and drinking water contaminants. The overall structure of ACToR is composed of a series of domains, linked together by chemicals.



Domain	Description	Location	
Chemicals	Structure, names and other basic chemical information	Main ACToR Database	
Assays	Quantitative and other tabular data on chemicals	Main ACToR Database	
Toxicology	In vivo study data from multiple domains	ACToR / ToxRefDB Database	
Genomics Microarray Data	Full microarray data sets, in both original and transformed versions	Under Development	
Biological Reference Data	Information on genes, proteins and pathways, downloaded from public sources	Main ACToR Database	
ToxMiner	Detailed data from the ToxCast and ToxRefDB programs - used for ToxCast analyses.	Separate ToxMiner database - linked to ACToR by chemical ID	

Currently, chemical toxicity data resides in a variety of specialized databases with incompatible formats and in many different locations. In the past, researchers needed to search many databases and manually organize the results in order to amass all the information on a given chemical. While this is possible for a few chemicals it is very difficult to compile comprehensive data sets on chemically-similar sets of compounds that can be explored with structure searching tools. Since the majority of chemicals in ACToR have chemical structures produced by the EPA ToxCast chemical prioritization program, ACToR will help to stimulate studies of structure-function relationships in sets of environmental chemicals. In addition, by bringing together data from a large number of sources and making the data structure-searchable, ACToR facilitates searches that transcend current databases in abundance of information and in number of generic chemicals.

The ACToR project is compiling data (both quantitative and qualitative) from a large number of sources (called <u>data collections</u>). Some of the sources include EPA, PubChem, other NIH, USDA and FDA databases; state, national, international and academic references. One novel data collection is <u>ToxRefDB</u> (Toxicology Reference Database), which includes detailed information on *in vivo* guideline study results for pesticides and other potentially toxic chemicals that has been assembled by the NCCT. ACToR is also the primary repository of data being produced by the EPA. Overall, ACToR is engineered to be flexible enough to add a variety of new data from sources with different formats in a straightforward manner. Because we have not already discovered and included all toxicology data, we are always interested in obtaining other data collections that could be incorporated into the system.

Overview: How to Use ACToR

In ACToR chemicals are organized into three main classes, the first two of which are modeled closely after the corresponding PubChem data model. The three main classes are:

- Substance- a unique chemical from a single "data collection"
- Compound- holds chemical structure information
- <u>Generic Chemical</u>- aggregates a chemical structure plus all of the corresponding substances. The common link is that all substances share the same CAS registry number.

Browsing

One can find information on ACToR information by either browsing or by searching for a particular chemical. A person can look through the ACToR database by browsing through either the data collections or the assay data.

- Assay: An assay is a collection of data for substances from one data collection.
- <u>Data collection:</u> A collection of chemical and assay data from a single source

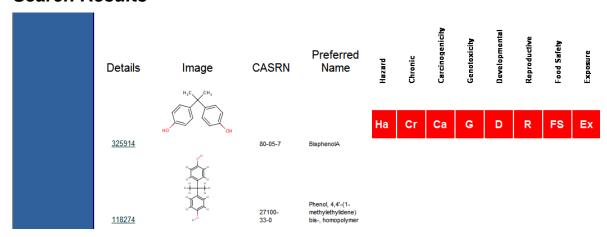
Searching

If a person has a particular chemical in mind, he or she can find it by either entering the name or CAS number or they can draw the structure.

- <u>Search by Name</u>- allows one to search for chemicals by their common and their many alternate names.
- Search by CAS- enables one to search by CAS number
- <u>Search by Structure</u>- allows a person to either draw a chemical or paste an INCHI or SMILES code onto the drawing pad to find chemicals.

After conducting a search, it is possible to go back and edit what has been entered. To do this, DO NOT CLICK ON THE BACK BUTTON. Clicking on the back button will erase all of the data. Instead re-click on the "Search by (name/CAS/structure)" link on the left navigation bar.

Search Results



The search results are composed of a table with all the chemicals in that data table. The columns include:

<u>Details-</u> contains the generic chemical id and a link to the chemical's generic chemical page

<u>Image</u> – Drawing of the chemical

CASRN- CAS Number

Preferred Name- the chemicals proper name

A series of red boxes will appear to the right of the "Preferred Name" column. The appearance of a red box indicates that there is toxicity data available for the substance that appears under the "Preferred Name" column. The initials within the box denote the specific type of toxicological information that is accessible with ACToR. It does not necessarily mean that the substance has tested positive for a particular effect.

<u>**H** (**Hazard**</u>)- This category includes substances that have been assessed for the potential to cause adverse health effects or damage to property.

<u>Ca (Carcinogenicity</u>)- This category includes substances that have been assessed for the potential to cause cancer.

<u>G (Genotoxicity</u>)- This category includes substances that have been assessed for the potential to cause damage to genetic material, such as DNA.

<u>**D** (**Developmental**</u>)- This category includes substances that have been assessed for the potential to affect the development and/or growth of an organism.

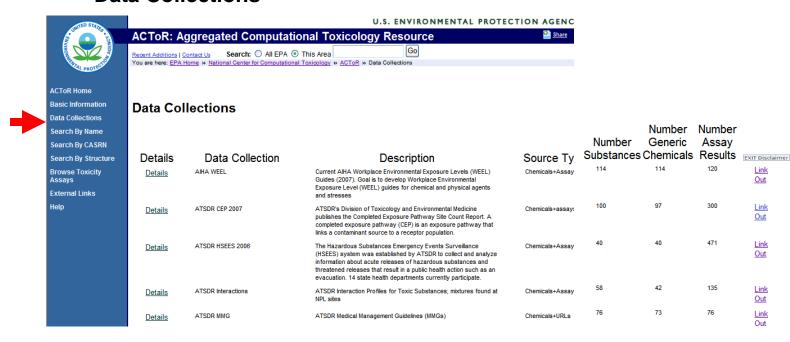
<u>R (Reproductive</u>)- This category includes substances that have been assessed for the potential to interfere with normal reproduction, such as causing changes in fertility.

<u>Cr (Chronic</u>)-Substances that have tested for adverse effects by exposure to the test agent over a substantial portion of the organism's lifespan.

FS (**Food Safety**)- Designation in this category means that the substance has been tested as to its safety as either: an additive to food, a food ingredient or present in food packaging.

Ex (Exposure)- Exposure information, including results of biomonitoring studies, concentrations in the air, water or soil, and allowable exposure limits.

Data Collections



All information ACToR is organized by <u>Data Collections</u>. A data collection often contain substances, chemical structures and <u>assays</u>. The entire list of data collections in ACToR can be seen by selecting the Data Collections in the left hand navigation bar. For each collection, the following data are presented:

Details- contains a link to individual data collection page

<u>Data collection</u>- a short synopsis of the data collection

Description- short synopsis of the data collection

Source type- the type of information contained in that data collection. It can include:

Chemical - chemical list of substances

 $\underline{Chemical + Assay}$ - a list of chemicals that have associated information $\underline{Chemical + Structure}$ - a list of chemicals and their corresponding structures $\underline{Chemicals + URL}$ - a list of chemicals having specific website for individual substances.

<u>EPA target list</u>- target list of chemicals to investigate created by the EPA <u>Search by CASRN and name</u>- chemical list with names and CAS numbers

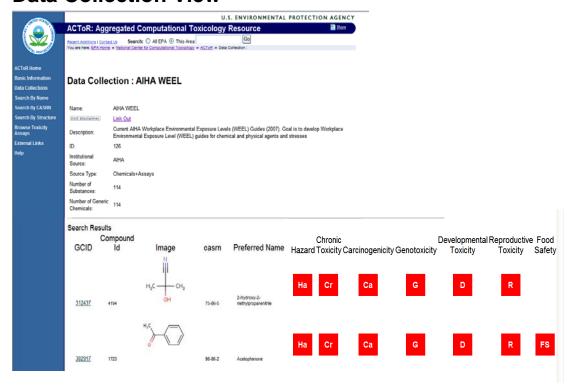
Number of substance- the number of substances within the data collection Number of generic chemicals the number of generic chemicals within the data collection. The number of generic chemicals may be less than number of substances if some substances do not have <u>CAS</u> numbers or if there are multiple substances with multiple names leaving identical CAS number.

<u>Number of assay results</u>- a measure of discrete assay for the substance contained in the collection.

Link out- provides a direct connection to the external website of a given data collection

To view the list of chemicals in the data collections, select the "Details" link at the left. This will take you to the <u>Data Collection View</u>. To navigate to the sources' external site select the "Link Out" hyperlink at the far right in data collections table.

Data Collection View



The data collection page shows a summary of the information contained within a specific data collection. This page can be reached by clicking on data collection page and then selecting one of the "Details" links. Here, the information is divided into three parts: overview, and search results.

Overview

The top chart provides a brief overview of the data collection. This includes:

Name - name of the data collection

<u>Link out</u> - provides a direct path to the source's external website

Description - description of the data collection

<u>ID</u> - the internal id number of the data collection used for database management purposes **Institutional Source** - the name of the institution that provided the data

<u>Source Type</u> – provides a cryptic description of the type of information contained in the data collection, e.g. a chemical list, an associated assay chemical- specific URLS, etc.

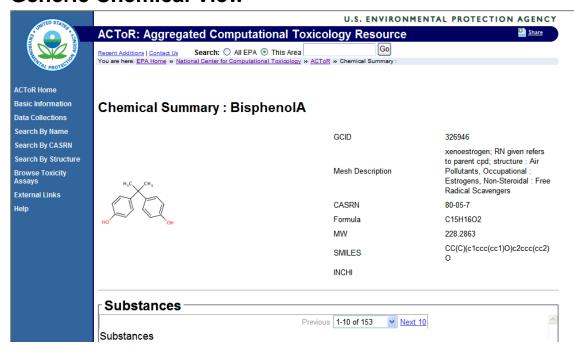
Number of Substances - the number substances in the data collection

<u>Number of Generic Chemicals</u> - the number of generic chemicals in the database. This number may be less than number of substance if some substances do not have <u>CAS</u> <u>numbers</u> or if there are multiple substances with the same CAS number.

Search Results

The search results give a list of all the chemicals that appear in the data collection. For more information go to "how to use ACTOR: search results".

Generic Chemical View



This page is accessed by clicking on "Details" for a specific chemical and provides links to all the information relating to this chemical i.e. from all data collections in the ACToR database. Data has been aggregated from all substances with a specific CASRN from all data collections. This page is divided into four sections: Chemical Summary, Substances, Synonyms and Toxicology Data.

Chemical Summary

<u>GCID</u>- Generic Chemical ID- an internal chemical ID within in ACToR <u>Mesh Description-</u> When available, a brief description of the chemical that came

from the MESH Database from the US Nation Library of Medicine

<u>CASRN</u>- The CAS (Chemical Abstracts Service) Registration Number

Formula- The chemical formula

MW- Molecular weight

<u>SMILES</u> - (<u>Simplified Molecular Input Line Entry System</u>) is a line notation used for representing molecules

Substance Data



The substance box provides a list of all the data collections that contain the chemical. Chemicals are aggregated by CASRN. The substance chart has the following categories:

Name- the chemicals proper name

<u>Data Collection</u>- the name of the data collection in which in appears

Synonyms

The synonyms box contains a list of all the <u>CAS</u> numbers and names used to describe the generic chemical.

Toxicology Data

The Chemical data is organized into different boxes. Each box has its own label and contains a list of <u>assays</u> and a summary of their results.

<u>Hazard</u> – Information on basic harm that can be caused by a chemical. Information under this category includes workplaces safety and first aid in case of exposure.

<u>Acute Toxicity</u> – Information on health effects due to short term exposure. <u>Subchronic Toxicity</u> - Information on health effects due to intermediate term exposure.

<u>Chronic Toxicity</u> - Information on health effects due to long term exposure. <u>Carcinogenicity</u> – Information from studies of cancer-causing ability of chemicals.

<u>Genetic Toxicity</u>- Information on the ability of chemicals to cause DNA damage. <u>Reproductive Toxicity</u>- Information on the ability of a chemical to damage an organism's reproductive ability.

<u>Neurotoxicity</u> - Information on the ability of a chemical to damage nerve cells or tissues.

<u>**Developmental Neurotoxicity**</u> – Information on chemicals that cause neurological deficits during development

<u>Immunotoxicity</u> - determines how a chemical affects the immune system <u>Dermal Toxicity</u>- includes studies about chemicals and what level of toxicity results from the substances being applied through the skin. May or may not have the skin as a target organ.

Respiratory Toxicity- contains data on how the chemicals that affect the respiratory system

Nephrotoxicity- measures to what degree that the chemical that affect the kidneys

Endocrine- contains data about if and how the chemical affect hormone signaling and downstream processes

Cardiotoxicity- contains data bout how the chemical affects the heart

Ecotoxicity- includes data about how chemicals affect non-human species such as fish and amphibians

<u>Food Safety</u>- includes data that determines if the food can be safely used as an ingredient, additive or food wrapper

<u>Toxicity other</u>- contains other information about a chemical's toxicity <u>PK/metabolism</u>- Information on pharmacokinetics and metabolism of xenobiotic chemicals

<u>Biochemical Manufacturing and Use Levels</u> – This is data compiled by the EPA on industrial chemicals that are subject to <u>TSCA</u> (Toxic Substances Control Act)

<u>Descriptive Data</u> – a variety of tabulated descriptive data, for instance on intended use of the compound

<u>Regulations to Which the Chemical is Subject</u> – This is a listing of international, U.S. Federal and state regulations and standards to which this chemical is subject.

<u>Material Safety Data Sheet</u> – if available, a link to the International Chemical Safety Card, which summarizes information from the Material Safety Data Sheet. <u>Physico-Chemical Data</u> – this is largely computed data on solubility, melting point, etc. based on chemical structure.

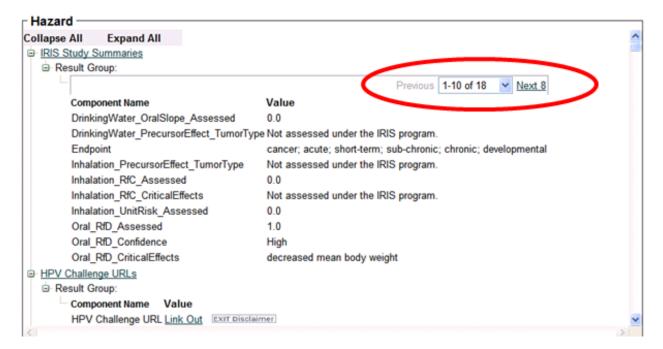
<u>Chemical Categories</u> – these are chemical structure categories used, for instance, to make initial predictions of toxicity of a new compound based on similarity with compounds which have already been tested.

<u>Pesticidal Mode of Action</u> – If the compound is a pesticidal active ingredient, this will provide the intended biological mode of action

<u>PubMed via MESH</u> – If available, a link to the literature in PubChem based on the MESH term.

Notes – notes provided on the chemical for and of the data collections.

<u>External Searches by NAME or CAS</u>- this is a list of pre-computed URL links to search external, on-line databases based on the CASRN and preferred name of the chemical



How to move through the chart

The assay names are green and underlined. Clicking on the assay names will take you to the assay view page. Clicking on the box next to the assay name will expand or collapse the result group. The result group has two columns:

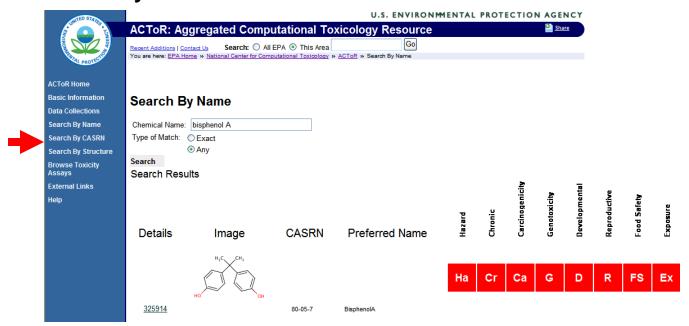
Component Name- the assay component name

Value- includes number, names, text and URLs.

Result Charts

If the chart contains more than 10 rows, then a list box and a "next __" link will appear in the upper right hand corner of the chart. To see all the data at once select the "Show all" option in the list box.

Search by Name



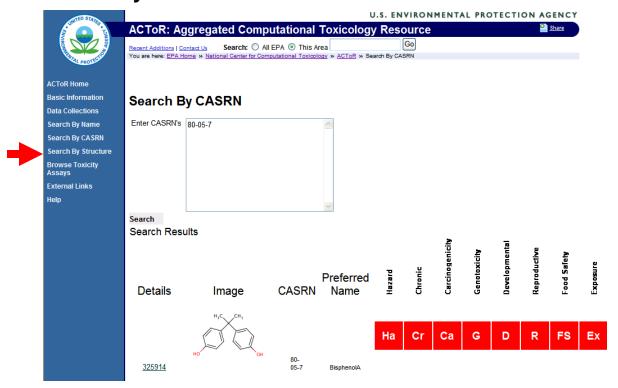
To search for a chemical, type in the full or partial name in the text box. Select either "exact match" or "any match". Exact match will find the chemical whose name matches what you typed in. "Any" match will find matches that are similar to what you typed in. The search is performed against all of the synonyms that have been compiled for each generic chemical. When one clicks the search button, a chemical list chart appears with the results. Note that the search by name program does not accept <u>SMILES</u> or <u>InCHI</u> notation. To use SMILES or InCHI see the <u>Search by Structure</u> page.

Search Results

See "How to use ACToR: Search Results".

After conducting a search, one can go back and edit what they entered. To do this, DO **NOT** CLICK ON THE BACK BUTTON. Clicking on the back button will erase all of your data. Instead re-click on the "Search by Name" button on the left navigation bar.

Search by CAS Number



Using <u>CAS numbers</u> is another way of locating and identifying chemicals. To find a chemical using a CAS number, type in one or more CAS numbers in the text box, separated by either commas or new lines. After search is completed, a standard chemical chart will appear.

Some examples of number in CAS format are:

7439-92-1

7440-50-8

79-34-5

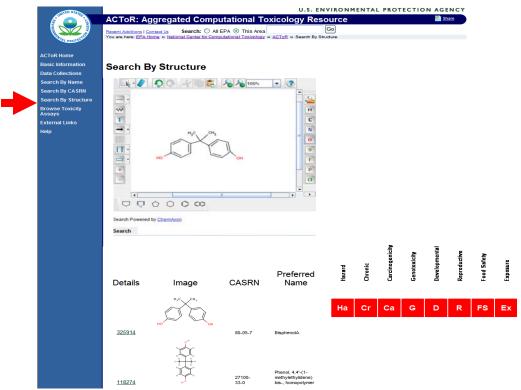
39001-02-0

Search Results

See "How to use ACToR: Search Results".

A search may be edited after being submitted. To do this, DO **NOT** CLICK ON THE BACK BUTTON. Clicking on the back button will erase all of your data. Instead re-click on the "Search by CAS" button on the left navigation bar.

Search by Structure



Drawing Board

There are two major ways to construct a molecule

- 1. The typical way to construct a molecule is to select a template (see 1), bonds (arrows 2), and atoms (see 3 and 4). First, select a template then click on the canvas (see 1). Then, click on button 2 and select the bond type. To attach a bond, place the cursor over the molecule until a purple circle appears. If there is a need to connect two molecules, then click and hold the left mouse button and drag the other end of the bond to the other molecule until another purple circle appears before letting go of the left mouse button. To add atoms, either click on one of the "quick add" buttons (3) or select button 4. Button 4 causes a small window to appear with the periodic table on it. Select an element and then click close. Clicking on the "Query" tab, gives some more options that do not appear on the periodic table.
- **2.** The fastest way to make molecules is to copy the molecule's <u>SMILES</u> or <u>InCHI string</u> and click on button 5.

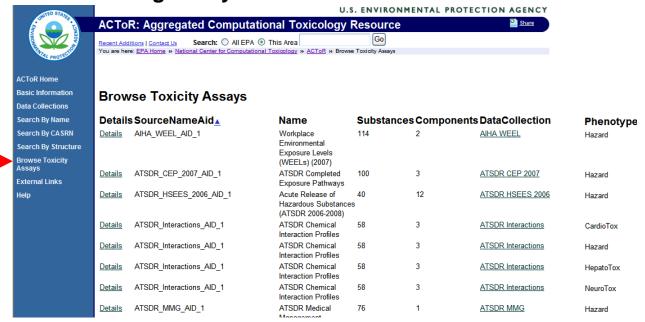
For a more in-depth tutorial for this program, click on button 6 on the upper right hand corner. This takes you to the ChemAxon Help Site.

Search Results

See "How to use ACToR: Search Results".

After conducting a search, one can go back and edit what they entered. To do this, DO **NOT** CLICK ON THE BACK BUTTON. Clicking on the back button will erase all of your data. Instead re-click on the "Search by Structure" button on the left navigation bar.

Browsing Assays



The browse assay page can be found by clicking on "Browse Toxicity Assay" in the left navigation button.

This page contains a collection of all the assays on the ACToR database. Assays are classified under one category but may have multiple phenotypes. Assays are organized alphabetically by assay name. Assays differ from data collections in that a data collection is all the information from a single source, while an assay is a collection of data arranged into a chart. So a single source can carry multiple assays.

<u>Details</u>- a link to a page that contains more details about a particular assay

Source Name AID- the location of the assay

Name- the name of the assay

Substances- the number of substances in the data

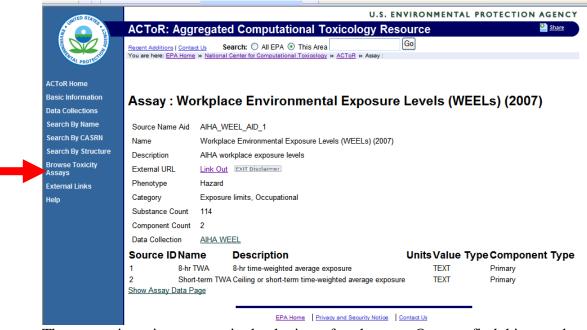
Components- the number of assay components

Data Collection- the name of the Data Collection

Category- The name of the Assays Categories

Phenotype- The name of the Assay Phenotypes

Assay View



The assay view gives a more in depth view of each assay. One can find this page by clicking on the "Details" link on the <u>browse assay page</u>. There are two sections to this page: the overview and the assay component chart.

Overview

Source Name AID- the location of the assay

Name- the name of the assay

Description- a brief summary of what the data entails

External URL- a link to the source

Components Count- the number of components

Data Collection- the name of the data collection

The above may be repeated if the assay has multiple phenotypes.

Assay Component Chart

Source ID Name		Description	Units Value Type Component Type		
1	8-hr TWA	8-hr time-weighted average exposure	TEXT	Primary	
2	Short-term TV	VA Ceiling or short-term time-weighted average exposure	TEXT	Primary	
Show Assay Data Page					

This section gives information about the type of data in the assay data. The assay component chart includes:

Source ID- the number assigned to the assay component within the assay

Name- the name of the assay component

<u>Description</u>- text description of the assay component

Units- what units are used for the assay component

Value Type- the type of values. Values include

Float- numbers with decimals. This includes standard decimal notation,

exponential notation and scientific notation.

Integer- whole numbers such as -2, 0, 1, 2

Categorical- contains categories such as colors

Boolean- true or false

URL- contains the URL from a hyperlink

<u>Text</u>- is in text form but is not a category, URL or True or False

Component type- tells whether this is a primary or a secondary value. The values include

Primary- a number that one would computer with

Modifier- includes Variance and SD

Variance- the distribution of a sample (Statistics)

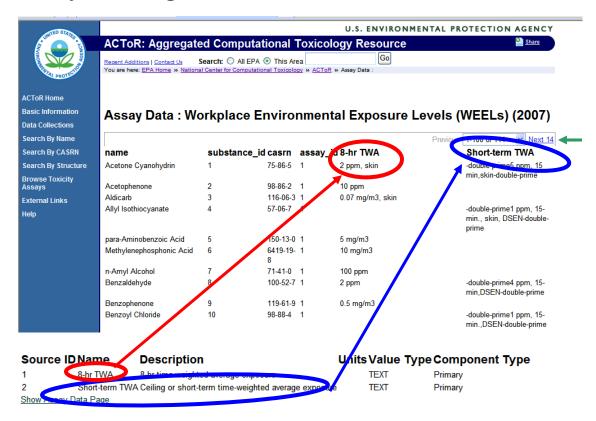
SD- standard deviation (Statistics)

Annotation- additional comments

Show Assay Data Page

This section contains a link that takes one to the Assay Data Page.

Assay Data Page



This page contains the <u>assay</u> data. In this chart, the columns headers are the <u>assay components</u> and the rows are the <u>substance</u>. The cells of this table are the <u>assay results</u>. The assay table can have more than one row or entry for the same substance, and elements in the data matrix can be empty.

One can find this page by clicking the link "Show Assay Data" on the Assay View page.

The charts vary with every assay. However, all of them contain the following:

Name- the name of the chemical

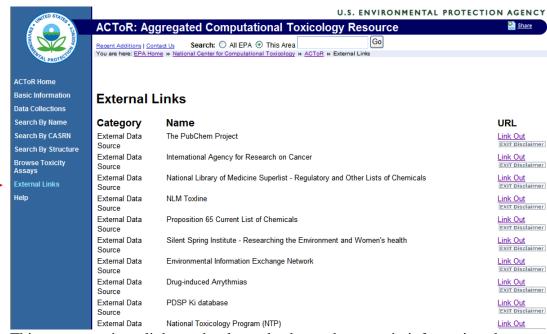
Substance ID- the substance ID number

Assay ID- the ID number for the assay within the data collection

The Remaining column headings are assay components. The component names can also be found on the Assay View Page's assay component chart.

Sometime, is the assay table is very long, there will be multiples "pages". One can learn how to navigate through the chart by going to here.

External Links



This page contains a link to other large databases that contain information about chemicals.

Category: are all external data sources

Name: name of the website URL: the hyperlink to the data

Glossary

Active ingredients

An active ingredient is a substance in a drug which has some pharmaceutical or pesticide values. This is the opposite for inactive ingredients, which are only carriers that allow that body to processes the active ingredients better.

ACTOR

(Aggregated Computational Toxicology Resource) is a collection of databases collated or developed by the EPA National Center for Computational Toxicology

Assay

An assay is a collection of data for substances from one data collection. Currently, an assay can be thought of a simple table. An assay falls into one data type category but may have multiple phenotypes. An assay can have more than one row or entry for the same substance, and elements in the data matrix can be empty.

Assay category

Assays are organized into a number of categories that describe the broad type of data presented. Several of these categories describe the level of biological organization being probed, while others describe the class of information being presented. The current sets of categories are:

<u>PhysicoChemical</u>- physical and chemical properties (in vitro and/or in silico) <u>Biochemical</u>- chemical processes in living organisms that are non-cell based <u>Genomics</u>- gene expression values or signatures

Cellular- cell-based assay

Tissue- tissue slice assay

Organ- focus on organs

Organism- focus on organisms (animal testing)

<u>In vivo toxicology (tabular primary)</u>- tabulated results from primary animal-based studies of chemical effects

<u>In vivo toxicology (study listing primary</u>)- primary studies are available but have not been tabulated

<u>In vivo toxicology (tabular secondary</u>)-tabulated data from secondary sources for in vivo toxicology studies

In vivo toxicology (summary calls)- derived summary determinations of risk In vivo toxicology (summary report via URL)- links to text reports on the web for which specific data values are not directly accessible in tabular form General Descriptive information- a brief description of the chemical Regulation- listings of chemicals that fall under specific environmental laws, government mandates, or standards

Chemical Category

A Chemical category is a listing of structural or use categories, often intended for prioritization efforts

<u>Chemical Summary URL</u>- link to chemical summaries

<u>Chemical Use Level</u>- the amount of chemicals produced or used

<u>Pesticidal mode of action (MoA)</u>- explains how to drug or pesticide works and what the chemical targets

Assay component

An assay component defines one column or element of an assay. A component has a unique ID, a name, a description, a data type, and optionally units.

Assay phenotype

Some assays are characterized by toxicology phenotypes. This allows one to organize the data in ACToR into broad toxicity areas. The current set of phenotypes are:

<u>Hazard</u> – Information on basic harm that can be caused by a chemical.

Information under this category includes workplaces safety and first aid in case of exposure.

<u>Acute Toxicity</u> – Information on health effects due to short term exposure.

<u>Subchronic Toxicity</u> - Information on health effects due to intermediate term exposure.

<u>Chronic Toxicity</u> - Information on health effects due to long term exposure.

<u>Carcinogenicity</u> – Information from studies of cancer-causing ability of chemicals.

Genetic Toxicity- Information on the ability of chemicals to cause DNA damage.

Reproductive Toxicity- Information on the ability of a chemical to damage an organism's reproductive ability.

<u>Neurotoxicity</u> - Information on the ability of a chemical to damage nerve cells or tissues.

<u>Developmental Neurotoxicity</u> – Information on chemicals that cause neurological deficits during development

<u>Immunotoxicity</u> - determines how a chemical affects the immune system <u>Dermal Toxicity</u>- includes studies about chemicals and what level of toxicity results from the substances being applied through the skin. May or may not have the skin as a target organ.

Respiratory Toxicity- contains data on how the chemicals that affect the respiratory system

<u>Nephrotoxicity</u>- measures to what degree that the chemical that affect the kidneys

Endocrine- contains data about if and how the chemical affect hormone signaling and downstream processes

<u>Cardiotoxicity-</u> contains data bout how the chemical affects the heart

Ecotoxicity- includes data about how chemicals affect non-human species such as fish and amphibians

<u>Food Safety</u>- includes data that determines if the food can be safely used as an ingredient, additive or food wrapper

<u>Toxicity other</u>- contains other information about a chemical's toxicity <u>PK/metabolism</u>- Information on pharmacokinetics and metabolism of xenobiotic chemicals

Assay result

An assay result is one data point for a single substance and a single assay component.

Assay types

There are two main types of assays: phenotypes and categories.

CAS

CAS (Chemical Abstract Services) Registry Number (for more information)

Some examples of number in CAS format are:

7439-92-1

7440-50-8

79-34-5

59325

39001-02-0

59001050

Chemical

A chemical is defined by a unique chemical ID in the database and can be either a substance or a compound.

Chemical structure

Diagram of a chemical- can be used to search for information about chemicals.

Compound

A compound is an entity with a chemical ID and chemical structure information, which may be a 2 or 3 dimensional molfile or a string representation. This can be SMILES or InCHII.

Data collection

A data collection is at minimum a set of substances with corresponding CAS registry numbers and names. Additional information may include chemical structures and assays. As mentioned above, a <u>generic chemical</u> links together data from many <u>data collections</u> on all <u>substances</u> that share a common <u>CAS</u> registry number. All data is initially compiled as part of a set of Data Collections.

Exposure data

Exposure studies, measure the amount of a substance that people and animals are exposed to. This data does not explain how the person was exposed to it or if it causes health problems.

Generic chemical

A generic chemical aggregates all data from all data collections for substances with a single given CAS number. It will have links to one or more substances and all of their related assay data, as well as all synonyms derived from the substances.

HPV and **MPV**

HPV stands for High Production Volume industrial chemicals and MPV for Medium Production Volume industrial chemicals

Includes the chemical structure and calculated physical chemical properties of compounds produced or imported into the United States

InCHI

The IUPAC International Chemical Identifier (InChI[™]) is a non-proprietary identifier for chemical substances (for more information).

Inert ingredient

An inert ingredient means any substance other than an active ingredient. Inert ingredients tend to be carriers for the active ingredients. (<u>for more information</u>)

In vitro

An experiment that is performed outside of a living organisms (for examples test tubes)

In vivo

Experimentation done on or inside of living organisms- other wise known as animal testing

MESH

is the U.S. National Library of Medicine's controlled vocabulary used for indexing articles for MEDLINE/ PubMed. MESH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts.

SMILES

SMILES (Simplified Molecular Input Line Entry System) is a *line notation* (a typographical method using printable characters) for entering and representing molecules and reactions. (<u>for more information</u>)

Substance

A substance is an entity with a chemical ID, one or more names (including a CAS number) and potentially a URL pointing to primary data. One special name for the substance is the "source name sid" which is a unique alphanumeric label from the source, which allows a unique link back to the source.